

DOCUMENT RESUME

ED 206 630

TM 810 380

AUTHOR Stalford, Charles B.
TITLE The Genesis of a Stakeholder Strategy for Evaluation Utilization.
PUB DATE 19 Oct 79
NOTE 11p.; Paper presented at the Annual Meeting of the Evaluation Research Society (Washington, DC, October 19, 1979).
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; Evaluators; Federal Programs; *Formative Evaluation; Program Evaluation; *Self Evaluation (Groups)
IDENTIFIERS Conflict of Interest; *Evaluation Utilization; National Institute of Education

ABSTRACT

Part of a forum on a Federal offices' search for a strategy to gain better evaluation utilization, this paper discusses the strategy's conception. Three categories of educational evaluations, defined by social and structural characteristics of the program environment, are identified: (1) trial evaluations, whose funding is not guaranteed beyond the test period; (2) ongoing program evaluation, which has an audience whose interests in the outcomes may conflict and; (3) continuing evaluation at the school district level. The focus is a paper by Stephen Weiner entitled: "Pathology in Institutional Structures for Evaluation and a Possible Cure." The major idea posited herein is that an evaluator's potential usefulness and ability to take issue with the aims of evaluation is constrained by organizational, political, and social factors and norms of legitimate evaluator behavior. Weiner's alternative strategy, subsequently utilized by the National Institute of Education in two evaluations, is presented. Based upon an analysis of weaknesses within the structure for government funded evaluations of social programs, the strategy substitutes formative evaluations (commissioned locally by individual programs and responsive to local stakeholder groups or those having vested interests in the outcomes) for the conventional third party evaluation structure. (ABF)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED206630

THE GENESIS OF A STAKEHOLDER STRATEGY
FOR EVALUATION UTILIZATION

Charles B. Stalford
Senior Associate, Team Leader for Evaluation
National Institute of Education

Prepared for Evaluation Research Society
Annual Meeting, October 19, 1979

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

☐ Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

C. Stalford

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 810 380

The Genesis of a Stakeholder Strategy for Evaluation Utilization

This forum is about one Federal office's search for a better strategy to gain evaluation utilization. People who have been involved in this strategy at various times and levels are here today. I will discuss the conception of the strategy. Norman Gold will describe its birth. Charles Murray of AIR will discuss its upbringing and Helen Branch of the Atlanta Schools will, in the fashion that schoolteachers sometimes will in discussing your child, tell you about how the strategy "really is" in practice.

The strategy had its origins in 1977. Then, Norman Gold and myself were part of a small coterie of evaluation specialists in a group at NIE (the Assessment of Innovative Developments Division, headed by Jeff Schiller) who were seriously dissatisfied with several of their experiences with evaluation utilization. Our personal and collective experiences there extended to evaluations of Head Start, Educational Performance Contracting and Experimental Schools, among others.

Our notion of utilization then was, I believe, fairly simple and at a "gut" level: we wanted to see some instrumental use of evaluation findings, either through funds allocations or visible changes in policy. I don't believe we were then attuned to

what I would call the "revisionist" school on evaluation utilization, namely that more utilization occurs than is commonly acknowledged and that the failure to perceive it is a failure to perceive the many forms such as symbolic and knowledge building, that utilization may take (Patton et al. 1978; Alkin, 1979). I agree with the "revisionist" thinking. Particularly for knowledge building or appreciative senses of evaluation to be assessed, some time has to elapse by definition for such processes to occur. That long term effects such as these may not always have been noticed may not have much to do, necessarily, with substantive aspects of the evaluation as they do the fleeting nature of people's involvement with anyone issue and the transitory nature of policy issues. Two years ago, for example, the present emphasis on minimum competency testing probably was not anticipated; while today it is high priority, two years from now it may not be.

Staff members in our group were not without some "successes" in utilization, although they are perceived to exist to varying degrees. Some of us, for example, were involved in the Head Start evaluation by Ohio-Westinghouse. While the merits of that study are still debated (Rein and White, 1978) it seems to have been used. Educational performance contracting was a clearer example: OEO's 1970-71 evaluation of the technique in which some of us were involved seemed to administer a coup de grace to the idea. Utilization, then, can take the form of decisionmakers deciding not to do something as well as deciding to do something.

A problem with the Head Start evaluation is one which still is before

us: how to structure an evaluation which will be acceptable to all of its various audiences. In the case of Head Start, no matter what uses were or were not made of the study by policy audiences, it was discarded as useless by the program management on several methodological grounds which had caused it to bitterly oppose the study when it was formulated.

It would be possible to categorize the world of educational evaluations into three broad groups which are defined not by methodology, but by the social and structural characteristics of the applicable program environment. Each category carries with it unique implications for evaluation utilization. They are:

- evaluation of experiments or demonstrations whose funding is not assured beyond the test period. These are "trial situations" in which a large body of practitioners may or may not be interested in the results of the evaluation. Perhaps the clearest example of such experimentation and evaluation is the negative income tax research. There, the programs did not exist except in prototype form and the outcome of the evaluation was primarily of interest to policymakers interested in the concept. To some extent, the evaluations of Cities in Schools and Push-Excel which we will discuss this morning are in the same category.

- evaluation of ongoing programs in which there is a program and practitioner audience interested in the outcomes as well as a policy audience. Such programs include the variety of discretionary ones funded by the state and Federal governments- anti-dropout, bilingual education, health centers, community mental health centers and the like. In these cases, the wishes of the evaluator, those commissioning the evaluation and those being evaluated may be in substantial conflict. In the Federal funding environment, this was the type of situation that I believe Chelmsky (1977) wrote about. It is also the one which the paper by Weiner et al. (1977) which I will discuss in a minute is very oriented to. The evaluator in this type case may well have to pick and choose among different audiences he is going to "please".

- evaluations at the school district level carried out on a continuing basis, with or without Federal or state funding support. While the literature on evaluation is largely made up of rather exciting if not always satisfying experiences, e.g., Follow Through, relatively little is known about the day to day practice of evaluations in schools. A recent study by the Center for the Study of Evaluation (Lyon et al., 1978) has shed a good deal of light on this area and we in NIE will be devoting more resources to it in the future. It is not the subject of this morning's meeting, so I will pursue it further. I do bring to your attention that the 1980 grants announcement of NIE's Teaching and Learning Program in the "Methodology" Section

contains an explicit invitation for proposals focused on evaluation at the local level. Further, we hope school districts themselves, possibly with SEA involvement will submit. This is something of a new precedent in this particular area for NIE and I hope you will publicize it to your friends who may be in school districts.

The focus of my introductory remarks this morning will be the 1977 paper by Dr. Stephen Weiner and associates. Dr. Weiner was then at the Stanford Evaluation Consortium; the Consortium had a contract with NIE to assist in an evaluation we were planning at the time. As an outgrowth of that work, Dr. Weiner and associates wrote a paper on the utilization of evaluation, which I have also made available this morning. It figured in our subsequent evaluation strategy for the Cities in Schools and Push-Excel evaluations which will be discussed subsequently.

The title of the paper "Pathology in Institutional Structures for Evaluation..." indicates important things about its content. These are (1) something is seriously wrong insofar as use of evaluations is concerned and (2) institutional, or what I would call structural aspects of the environment for evaluations is significantly determining people's behavior in this regard. Weiner et al. do not set aside questions of quality or timing of evaluative information as contributing to utilization. They do, however, take a quite different tact in their institutional focus than do Patton et al. (1978) in the latter's focus on "personal qualities" which contribute to utilization.

Neither Patton nor Weiner would, I believe, assert that their approaches are mutually exclusive. Whether one wishes to see "personal qualities" or "institutional factors" as important in utilization may be in part a factor of one's predisposition to psychology or sociology in explaining behavior. Insofar as the institutional, or sociological side of the explanation is concerned, I have written elsewhere (1980, in press) that social structures in evaluation may not so much "shape" behavior in deterministic fashion so much as facilitate (or impede) various forms of behavior.

Citing Caplan et al. at Michigan (1975), Weiner draws the following central conclusion regarding evaluation utilization:

The main factors which appear to limit the level of utilization can be found in that portion of the theoretical literature on utilization which can be characterized as the Two Communities Theories. The essential line in this theory is that the main reasons for the nonuse of knowledge can be understood by examining the relationships of the researcher and the knowledge production process to the policymaker and the policymaking process. More specifically, it suggests that social scientists and policymakers operate in separate worlds with different and often conflicting values, different reward systems, and different languages. Our data suggest that mutual mistrust is an important factor in the separation of the producer and user communities.

Weiner suggests a strategy "based upon an analysis of weaknesses inherent in the existing institutional structure for government funded external evaluations of social programs". He goes on:

Consequently our suggestions call for structural modifications intended to overcome these weaknesses. Principally, we argue that existing linkages between funding agencies and evaluation contractors should be substantially attenuated to relieve contractors of the requirement to conform to norms of appropriate activity imposed by funding agencies. We recommend,

instead, various devices which would draw evaluation contractors into an interaction with those groups that direct interest in the kind and quality of the services delivered in the programs being evaluated. We suggest further that 'knowledge aggregation and transfer' groups, specializing in particular issue areas, be established. These groups would take responsibility for developing a broad view of public programs in their area, with a primary commitment to shaping public policy discussion based upon analysis of program successes and failures over time.

Esseentially, Weiner recommended dropping the conventional third party evaluation structure. In its place he would substitute formative evaluations commissioned locally by individual programs. These studies would be largely responsive to local "stakeholder" groups, or those who have an interest in the services being provided and the outcome of the evaluations.

The "knowledge aggregation and transfer groups" would act as something of a supreme court, hearing and considering evidence from program evaluation and other sources on the effectiveness of individual programs and providing long term judgements on success and failure in broad program areas rather than on the merits of individual programs.

What is of particular interest here because it is the concept which NIE picked up in its own thinking is that of "stakeholder" groups. Norm will discuss that in more detail. I will only briefly describe some of the indices of "pathology" that Weiner saw in the typical interactions among funding agencies, program managers and evaluators.

Weiner et al. see evaluators' potential usefulness being constrained by organizational and political factors and also norms of legitimate behavior expected of them as experts. ⁴ Organizationally, evaluators are seen to be overly dependent on needs of funding agencies out of concern for such things as future business. The funding agencies themselves are viewed as being excessively motivated in sponsoring evaluations by such mundane concerns as organizational maintenance and avoidance of controversy. Program managers are, of course, fearful of evaluations.

Politically, Weiner et al. see evaluators as not being free to take issue with basic aims or features of evaluations once commissioned and hence not able to contribute to broad issues raised by such studies to the extent that their intellectual capacities would permit. Finally, their norms of legitimate behavior are seen to constrain evaluators from raising controversial issues of potential interest to program clients and other groups affected by an evaluation, but which the funding agency might just as soon sidestep.

One might characterize the environment which results from this confluence of factors as being unsatisfactory to everyone concerned. Most of us will suspect, recognize some of our own experiences in it. In another program evaluation setting, we at NIE did something different with the structure which was designed to deal with some of the "institutional pathologies" which Weiner identified (Stalford, 1980, in press). In that case, rather than fund a traditional third party evaluation of one of our programs, we transferred the funds to the program and told them to commission the evaluation themselves! We judge the results to be generally satisfactory

REFERENCES

- M. Alkin, P. Daillak and P. White. Using Evaluations. Does Evaluation Make a Difference? Beverly Hills: SAGE Publications. 1979.
- N. Caplan et al. The Use of Social Science Knowledge in Policy Decisions at the National Level. Ann Arbor: Institute for Social Research, University of Michigan, 1975.
- E. Chelmsky. A Symposium on the Use of Evaluation by Federal Agencies. Mitre Corporation. July 1977.
- C. Lyon et al. Evaluation and Decision-making in School Districts. Los Angeles: Center for the Study of Evaluation, UCLA. 1978.
- M. Patton et al. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research" in T. Cook (Ed.) Evaluation Studies Review Annual, Vol. 3. Beverly Hills: SAGE Publications, 1978.
- M. Rein and S. White. "Can Policy Research Help Policy?" in Evaluation Studies Review Annual, Vol. 3.
- C. Stalford. "Innovation in Third Party Evaluation Structures: The NCHEMS Case" in L. Braskamp and R. Brown (Eds.) Utilization of Knowledge and Evaluative Information. Jossey Bass, in press.
- S. Weiner et al. "Pathology in Institutional Structures for Evaluation and a Possible Cure". Stanford Evaluation Consortium (unpublished). August 1977.